Clinical Journal of Sport Medicine

Genomics and biology of exercise, where are we now?

--Manuscript Draft--

Manuscript Number:	
Full Title:	Genomics and biology of exercise, where are we now?
Article Type:	Editorial
Corresponding Author:	Nir Eynon, PhD Victoria University Melbourne, Victoria AUSTRALIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Victoria University
Corresponding Author's Secondary Institution:	
First Author:	Danielle Hiam
First Author Secondary Information:	
Order of Authors:	Danielle Hiam
	Patrice Jones
	Yannis Pitsiladis
	Nir Eynon
Order of Authors Secondary Information:	
Manuscript Region of Origin:	AUSTRALIA

Editorial

Genomics and biology of exercise, where are we now?

Danielle Hiam[1,2], Patrice Jones[1], Yannis Pitsiladis[3], Nir Eynon[1]

Affiliations:
[1] Institute for Health and Sport (iHeS), Victoria University, Melbourne, Australia.
[2] Deakin University, Geelong, Australia, Institute for Physical Activity and Nutrition (IPAN), School of Exercise and Nutrition Sciences.
[3] Centre for Sport and Exercise Science and Medicine, University of Brighton, United Kingdom.

Corresponding author:
Professor Nir Eynon
Institute for Health and Sport (iHeS),
Victoria University
PO Box 14428
Melbourne, VIC 8001, Australia.
Tel: (61-3) 9919 5615, Fax: (61-3) 9919 5532, E-mail: Nir.Eynon@vu.edu.au

Genomics and the biology of exercise is an exciting research field aiming to understand how genetic variations interact with the environment to facilitate adaptations to exercise in healthy, and diseased, as well as elite athlete populations [1]. Adaptations to exercise interventions, elite athletic performance and predisposition to sports injuries are complex, polygenic traits and seems to be highly heritable [2]. This has stemmed from seminal studies like the HERITAGE family study (HEalth, RIsk, exercise training And GEnetics) [3] which initiated a wealth of literature investigating candidate genetic variants influencing exercise adaptations and sport performance [4] including the well characterised ACE Insertion/Deletion [5], and the ACTN3 polymorphisms [6, 7].

With the establishment of international consortia such as the HapMap project, 1000 Genomes Project and Noncoding Human Genome Sequence (ENCODE) project in the 2000s, it became evident that appropriate powered studies with large sample size were the new way forward in the understanding of biological insights in complex traits. Recent years has seen advances in the technology and affordability of large-scale data analysis, termed "omics". This has resulted in the field moving beyond small candidate gene driven analysis to an unbiased, hypothesis-free genome wide association analysis (GWAS). One such initiative was the GAMES consortium which conducted a GWAS to identify potential genetic variants associated with world class endurance athletes [8]. This was followed by 16 loci identified with grip strength which a proxy of muscular fitness conducted in 195,180 individuals from the UK Biobank cohort [9]. It became clear that complex traits, like those associated with exercise adaptation and performance, are influenced by thousands of genetic variants with small effect sizes.

National and international collaborative research efforts are key to overcome the limited sample size and allow access to diverse cohorts of different ancestry. While there is no database that is near extensive enough for sport and exercise medicine, there are concerted efforts to build a large sample multi-centred database in sport and exercise medicine. These include the Athlome consortium project (www.athlomeconsortium.org) established in 2015 and consisting of multiple cohorts including the GeneSMART cohort [10], ELITE (http://elite.stanford.edu/), POWERGENE, GENESIS, GOINg and the Netherlands Twin Register cohort [11]. These and other initiatives will allow for population-based approaches to understand the role of genetic and environmental factors contributing to the complex exercise response phenotypes [12].

Replication and validation are critical. The vast majority of genes elucidated have not been meticulously replicated in independent cohorts. This is critically important as it increases the likelihood that the results are biologically relevant and reduces the number of false positives [12]. This is often omitted in sports and exercise medicine and has resulted in a "messy" understanding in the role of the genome in exercise adaptations and performance with a large number of false positives. Further, the overall evidence from the literature connecting causal genes to exercise response and performance is relatively low. This poses a problem given the expanding interest in direct-to-consumer genetic testing services, which promise personalised training programs and information on one's athletic potential, despite a lack of reproducible and supporting evidence [13-15]. If we hope to understand causal variants or genes it is vital we begin to integrate "omic" technologies from genome and epigenome to transcriptome and proteome and metabolome so we can capture a complete picture of complex human traits such as exercise response and performance.

Moving beyond the genome. Since the completion of the ENCODE project [16] the field is now moving into integrating the genome with other levels of regulation of protein coding genes such as the epigenome [17], transcriptome [18] and proteome [19] to gain a broader understanding on how harbouring a set of genetic variants have biological effects on exercise adaptations. Professor Bouchard, a pioneer of this field of research summarised this well "*In retrospect, exercise genomics research paradigms were naïve in the early days as it was assumed that there was more or less direct and unequivocal relations between DNA sequence variants and traits of interest. This simple model did not recognise the highly complex distributed regulation of gene expression and of post transcription and post-translation events*" [1].

Certainly, in recent years literature has come to light illustrating the importance of post transcription and post-translational regulators in programming sport performance and exercise responses. With this comes new challenges with specialised bioinformation's and biostatistical data scientists to process and integrate the huge datasets and bench-top scientists to validate these findings. For example, expression quantitative trait loci (eQTLs) analysis leverages gene loci identified from GWAS and integrates these with the transcriptome to identify differential gene expression levels and uncover the 'molecular phenotype' leading to these variations in exercise response [20, 21].

What does the future hold? Now more than ever there is a need for multi-disciplinary teams and world-wide collaborations if we wish to make significant and impactful advances in the field [22]. The COVID-19 pandemic highlighted the utility of developing such collaborations and harnessing "big data", to ensure growth in the field during changing times. Collaborative efforts to untangle the complex interactions between genes and the environment are critical in determining predisposition to high-level performance, fitness, or disease. "Omic"-based approaches hold further promise in advancing gene doping detection methods in sport [23, 24]. Progress in the field of exercise genomics will require the development of field-specific resources, which may draw inspiration from existing platforms such PhenomeXcan [25] and the GWAS Catalog [26], that enable the simple exploration and cataloguing of gene- and SNP-trait associations. These advances are expected to make genetic testing for exercise prescription a greater reality. This will inevitably introduce ethical and confidentiality issues that will need to be addressed on an ongoing basis. It will be essential that scientists and others work in line with ethical guidelines developed by the collaborative efforts of The International Federation of Sports Medicine (FIMS) and Athlome Project Consortium, to enable progress whilst preventing the misuse of "omic" data [13-15]. The management and storage of "omic" data will require further modern technologies, such as blockchain [27, 28], which will provide new possibilities for data security and integrity in exercise genomics.

In the field of exercise genomics & exercise biology, the utmost question is not how genetics may be used to identify talents, but how the knowledge derived from genetic research may help to stimulate the potential of athletes, prevent injuries, and help the recovery of athletes from injuries, as well as how the knowledge may contribute to the understanding of public health. We are now at precipice where understanding the genome in isolation is no longer a viable option, and we need to integrate the human genome with other levels of regulation. In this special edition of exercise genomics, we highlight some of the recent work coming from the Athlome consortium and FIMS members. This includes studies assessing not only at the genome but also the transcriptome and epigenome (microRNA's) of elite athletes and responses to exercise.

References

1. Bouchard, C., *Exercise genomics--a paradigm shift is needed: a commentary.* Br J Sports Med, 2015. **49**(23): p. 1492-6.
2. Bouchard, C., et al., *Familial aggregation of VO(2max) response to exercise training: results from the HERITAGE Family Study.* J Appl Physiol (1985), 1999. **87**(3): p. 1003-8.
3. Bouchard, C., et al., *The HERITAGE family study. Aims, design, and measurement protocol.* Med Sci Sports Exerc, 1995. **27**(5): p. 721-9.
4. Williams, C.J., et al., *Genes to predict VO2max trainability: a systematic review.* BMC Genomics, 2017. **18**(Suppl 8): p. 831.
5. Montgomery, H., et al., *Angiotensin-converting-enzyme gene insertion/deletion polymorphism and response to physical training.* Lancet, 1999. **353**(9152): p. 541-5.
6. Yang, N., et al., *ACTN3 Genotype Is Associated with Human Elite Athletic Performance.* The American Journal of Human Genetics, 2003. **73**(3): p. 627-631.
7. Houweling, P.J., et al., *Is evolutionary loss our gain? The role of ACTN3 p.Arg577Ter (R577X) genotype in athletic performance, ageing, and disease.* Hum Mutat, 2018. **39**(12): p. 1774-1787.
8. Rankinen, T., et al., *No Evidence of a Common DNA Variant Profile Specific to World Class Endurance Athletes.* PLoS One, 2016. **11**(1): p. e0147330.
9. Willems, S.M., et al., *Large-scale GWAS identifies multiple loci for hand grip strength providing biological insights into muscular fitness.* Nat Commun, 2017. **8**: p. 16015.
10. Yan, X., et al., *The gene SMART study: method, study design, and preliminary findings.* BMC Genomics, 2017. **18**(Suppl 8): p. 821.
11. Pitsiladis, Y.P., et al., *Athlome Project Consortium: a concerted effort to discover genomic and other "omic" markers of athletic performance.* Physiol Genomics, 2016. **48**(3): p. 183-90.
12. Alvarez-Romero, J., et al., *Mapping Robust Genetic Variants Associated with Exercise Responses.* Int J Sports Med, 2021. **42**(1): p. 3-18.
13. Vlahovich, N., et al., *Genetic testing for exercise prescription and injury prevention: AIS-Athlome consortium-FIMS joint statement.* BMC Genomics, 2017. **18**(Suppl 8): p. 818.
14. Tanisawa, K., et al., *Sport and exercise genomics: the FIMS 2019 consensus statement update.* Br J Sports Med, 2020. **54**(16): p. 969-975.
15. Webborn, N., et al., *Direct-to-consumer genetic testing for predicting sports performance and talent identification: Consensus statement.* Br J Sports Med, 2015. **49**(23): p. 1486-91.
16. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
17. Jacques, M., et al., *Epigenetic changes in healthy human skeletal muscle following exercise– a systematic review* Epigenetics 2019.
18. Deane, C.S., et al., *Transcriptomic meta-analysis of disuse muscle atrophy vs. resistance exercise-induced hypertrophy in young and older humans.* J Cachexia Sarcopenia Muscle, 2021.
19. Molendijk, J. and B.L. Parker, *Proteome-wide Systems Genetics to Identify Functional Regulators of Complex Traits.* Cell Syst, 2021. **12**(1): p. 5-22.
20. Nicolae, D.L., et al., *Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS.* PLoS Genet, 2010. **6**(4): p. e1000888.
21. Keildson, S., et al., *Expression of phosphofructokinase in skeletal muscle is influenced by genetic variation and associated with insulin sensitivity.* Diabetes, 2014. **63**(3): p. 1154-65.
22. Tanaka, M., G. Wang, and Y.P. Pitsiladis, *Advancing sports and exercise genomics: moving from hypothesis-driven single study approaches to large multi-omics collaborative science.* Physiol Genomics, 2016. **48**(3): p. 173-4.
23. Durussel, J., et al., *Blood transcriptional signature of recombinant human erythropoietin administration and implications for antidoping strategies.* Physiological Genomics, 2016. **48**(3): p. 202-209.

24. Wang, G., et al., *Validation of whole-blood transcriptome signature during microdose recombinant human erythropoietin (rHuEpo) administration.* BMC Genomics, 2017. **18**(8): p. 817.
25. Pividori, M., et al., *PhenomeXcan: Mapping the genome to the phenome through the transcriptome.* Science Advances, 2020. **6**(37): p. eaba2083.
26. Buniello, A., et al., *The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019.* Nucleic Acids Research, 2019. **47**(D1): p. D1005-D1012.
27. Jin, X.-L., et al., *Application of a Blockchain Platform to Manage and Secure Personal Genomic Data: A Case Study of LifeCODE.ai in China.* Journal of medical Internet research, 2019. **21**(9): p. e13587-e13587.
28. Ozercan, H.I., et al., *Realizing the potential of blockchain technologies in genomics.* Genome research, 2018. **28**(9): p. 1255-1263.